



The Commission's Student Data System: A Progress Report

Working Paper W/P/05-07 December 2005 • www.cpec.ca.gov

In May 2005, the University of California, the California State University, and the California Community Colleges Chancellor's Office provided the Commission with enrollment and degree data that identifies individual students. The data was provided in accordance with AB 1570, which authorizes the Commission to use a student identifier in its data system in order to conduct longitudinal studies of the progress of individual students. This data will enable the Commission to conduct analysis of issues such as the time taken for students to complete their degrees, the success rate of students transferring from community colleges to the four-year systems, and success and persistence of students as they progress through the systems.

Because social security numbers are used to identify students in the data files, it is of utmost importance that the data be handled in a way that protects its confidentiality. The Commission's information technology staff outlined a series of security procedures in its paper *Implementing the Commission's Student Record System—First Steps*, presented at the June 2005 Commission meeting. Over the past six months, staff has developed the specifics of these procedures, acquired equipment, and written the computer programs needed to process the data. The original data files are processed on computers that are not connected to the Commission's network. In the data files available to research staff, social security numbers are replaced with an identification number assigned in the first step of the process. The procedures ensure that this identification number cannot be traced back to a social security number.

Data now available for analysis

The data files on degrees awarded are now available for use by Commission research staff. The university systems provided data for 2000 through 2004 and the community colleges provided data for 1993 through 2004. The data consisted of nearly two million records.

Most of the records were originally identified by social security numbers. For example, the 2003 data for CSU had records for 79,986 degrees and certificates awarded. Because some students were awarded

Identification codes in 2003 degrees file

	University of California	California State University	Community Colleges
Total records	52,549	79,986	118,405
No ID given	10,304	none	none
Distinct IDs	41,539	78,929	96,743
Issued SSN	41,401	77,176	94,027
Un-issued SSN	15	983	106
College-issued ID	123	770	2,610

Issued SSNs are those whose first five digits correspond to a combination of area code and group code issued by the Social Security Administration.

Un-issued SSNs are those with an invalid combination of area code and group code in the first five digits

more than one CSU degree in 2003, there were 78,929 distinct student IDs in the 2003 data. Of these, 78,159 were social security numbers. A check against the Social Security Administration's list of issued prefixes showed that 983 of the SSNs could not have been issued, leaving 77,176 apparently valid SSNs. The data also includes an additional 770 ID codes issued by CSU campuses to students lacking social security numbers.

The data from UC was less well identified. Over 10,000 degree records for 2003 had missing IDs. However, nearly all of the remaining records had IDs originating from issued SSNs.

Use of the data

Because most of the data originates from records with actual social security numbers rather than college-issued ID codes, it is possible to track students from system to system. The enrollment data will be needed before staff can investigate issues such as the progress and persistence of students, but a limited analysis can be done with just the degrees data.

One issue of interest is “reverse transfer”. There are concerns that university graduates are enrolling at community colleges in order to get marketable job skills. However, the data shows that this is not a major trend. The table below shows the number of UC and CSU graduates who earned bachelor’s degrees in 2000 and earned a community college degree or certificate between 2001 and 2004. Relatively few university graduates show up in later years as community college graduates.

The degree data includes only those students who actually completed a community college degree or certificate within three years of earning a bachelor’s degree. Enrollment data would give a more

University graduates later earning a community college degree or certificate

	Graduates earning bachelor’s degrees in 2000*	Number earning a community college award, 2001–2004	
		Associate	Certificate
UC Berkeley	4,425	13	9
UC Davis	3,409	13	9
UC Irvine	2,499	24	2
UC Los Angeles	4,748	10	10
UC Riverside	1,514	10	9
UC San Diego	2,992	5	10
UC Santa Barbara	3,306	22	9
UC Santa Cruz	1,617	8	4
CSU Bakersfield	326	–	2
CSU Chico	1,018	1	15
CSU Dominguez Hills	952	–	2
CSU Fresno	1,183	6	16
CSU Fullerton	2,180	13	15
CSU East Bay	938	7	6
Humboldt State University	541	5	4
CSU Long Beach	2,278	13	10
CSU Los Angeles	1,173	5	12
CSU Northridge	1,980	18	5
Cal Poly, Pomona	1,018	1	2
CSU Sacramento	2,139	22	13
CSU San Bernardino	635	4	3
San Diego State University	2,499	24	11
San Francisco State University	1,827	17	9
San José State University	2,028	17	9
Cal Poly, San Luis Obispo	1,321	6	10
Sonoma State University	526	3	5
CSU Stanislaus	446	13	4
CSU San Marcos	731	8	2
CSU Monterey Bay	135	1	–

*Graduates identified by SSNs in the original data files

complete picture, showing the number of UC and CSU graduates enrolled at community colleges. However, the enrollment data files are larger and more complex than the degree data files and processing is not yet complete. When the enrollment data is available, staff will be able to produce summaries of the number of units taken at community colleges by UC and CSU graduates.

The table to the left is only one way of summarizing the data. Much more detail is available. For example, the same data can be summarized by discipline, as well as campus, to show the number of graduates in a particular discipline who enroll at community colleges.

The student data can be applied to a wide variety of other issues. It could be used to identify how persistence and success varies with students’ ethnicity and characteristics of their originating high schools. Another application would be to look at time-to-degree and units-to-degree for transfer students. This analysis would show whether transfer is really providing the benefits that should be expected from allowing students to complete their lower division studies at a local, lower-cost system.